Message Passing Algorithm for Distributed Downlink Regularized Zero-forcing Beamforming with Cooperative Base Stations

Chao-Kai Wen*, Jung-Chieh Chen†, Kai-Kit Wong‡, and Pangan Ting§

Abstract

Base station (BS) cooperation can turn unwanted interference to useful signal energy for enhancing system performance. In the cooperative downlink, zero-forcing beamforming (ZFBF) with a simple scheduler is well known to obtain nearly the performance of the capacity-achieving dirty-paper coding. However, the centralized ZFBF approach is prohibitively complex as the network size grows. In this paper, we devise message passing algorithms for realizing the regularized ZFBF (RZFBF) in a distributed manner using belief propagation. In the proposed methods, the overall computational cost is decomposed into many smaller computation tasks carried out by groups of neighboring BSs and communications is only required between neighboring BSs. More importantly, some exchanged messages can be computed based on channel statistics rather than instantaneous channel state information, leading to significant reduction in computational complexity. Simulation results demonstrate that the proposed algorithms converge quickly to the exact RZFBF and much faster compared to conventional methods.

Index Terms—Base station cooperation, Belief-propagation, Distributed algorithm, Message passing, Zero-forcing beamforming.

*Institute of Communications Engineering, National Sun Yat-sen University, Taiwan. Email: chaokai.wen@mail.nsysu.edu.tw.

†The Department of Optoelectronics & Communication Engineering, National Kaohsiung Normal University, Kaohsiung, Taiwan.

‡Department of Electronic and Electrical Engineering, University College London, UK.

§Industrial Technology Research Institute (ITRI), Hsinchu 310, Taiwan, R.O.C.

I. Introduction

Multiuser multiple-input multiple-output (MU-MIMO) antenna system has been recognized as an effective means to increase capacity in the downlink [1–3]. However, MU-MIMO may not be as effective if edge-of-cell users are concerned due to the severe inter-cell interference that is hard to suppress. In recent years, it has emerged that letting base stations (BS) cooperate can greatly improve the link quality of the edge-of-cell users by turning unwanted interference into useful signal energy, e.g., [4–9] (and the references therein). Ideally, by sharing all the required information via high-speed backhaul links, all BSs in a downlink cellular network can become a super BS with distributed sets of antennas. This architecture will then allow the use of well-known optimal or suboptimal transmission strategies such as capacity-achieving dirty-paper coding (DPC) techniques [5, 6, 10] and zero-forcing beamforming (ZFBF) [7, 11], respectively.

Although DPC is capacity-achieving, it is very complex and massive interest has been to employ ZFBF with a simple scheduler to approach near-capacity performance [7, 11]. For example, several testbeds for implementing BS cooperation have adopted ZFBF techniques, e.g., [12–15]. Regularized ZFBF (RZFBF) is a generalization of ZFBF by introducing the regularization parameter [16, 17]. It has been revealed that several beamformers can have a RZFBF structure by selecting the regularization parameter properly [18]. Even though information-theoretic studies have provided overwhelming support to RZFBF [18–20], the real question is how could RZFBF be implemented in a very large-scale cellular network?

A straightforward way to implement RZFBF would be to require that there is a central processing unit which possesses all the necessary channel state information (CSI) and performs the entire optimization. However, as a network expands with more BSs cooperating, it becomes inviable to perform joint processing over all BSs because of the limiting backhaul capacity and the excessive computational complexity. It is therefore of greater interest to consider an architecture where BSs only communicate with neighboring BSs and the overall computation cost is decomposed into many smaller computational tasks, amortized by groups of smaller number of cooperating BSs. Motivated by this, in this paper, we propose two message passing algorithms to realize RZFBF in a distributed manner. The proposed approaches are particularly well suited to cooperation of large clusters of simple and loosely connected BSs. Most importantly, in our designs, each BS is only required to know the data symbols of users within its reception range rather the entire cellular network, greatly reducing the backhaul requirements.

The use of distributed methods in beamforming computations has been studied recently in [21–26]. Our approach is similar to [21] in that both aim at achieving RZFBF and use belief propagation (BP). Nonetheless, the two approaches differ considerably. Our main contributions are summarized as follows:

- First, we generalize the earlier results in [21] to incorporate *multiple antennas* at both BSs and user equipments (UEs) and our results can be applied to a wide range of scenarios with *complex-valued systems*. Further, we adopt the approximate message passing (AMP) method in [27] to significantly reduce the number of exchange messages. The proposed AMP-RZFBF exhibits the advantage that every communication of BS with its neighbors only takes place in a broadcast fashion as opposed to the unicast manner in [21]. The used AMP method has recently received considerable interest in the field of compressed sensing [27–30]. Our form of the message passing algorithm is closely related to the AMP methods in [29] which are a special case of the generalized AMP [30].

- In AMP-RZFBF, BSs must compute several matrix inversions for every channel realization and then exchange these auxiliary parameters among themselves, requiring very high computational capability and rapid information exchange between the BSs. To tackle this, we approximate some of the auxiliary parameters by exploiting the spatial channel covariance information (CCoI). The CCoI-aided AMP-RZFBF results in significantly simpler implementations in terms of computation and communication. With the CCoI-aided AMP-RZFBF, the BSs compute and exchange the auxiliary parameters at the time scale merely at which the CCoI changes but not the instantaneous CSI. Simulation results show that CCoI-aided AMP-RZFBF achieves promising results, which are different from earlier results based on the CCoI, e.g., [31, 32], where a performance degeneration is usually expected.

- Implementing RZFBF in a distributed manner can be achieved by an optimization technique called the alternating direction method of multipliers (ADMM) approach in [33]. Applications of ADMM to the concerned beamforming problem can be found in [34] (or [33, Section 8.3]). However, it is known that ADMM can be very slow to converge. Simulation results will demonstrate that our proposed message passing algorithms exhibit a much faster convergence rate when compared to ADMM.

Notations—Throughout this paper, the complex number field is denoted by $\mathbb{C}$. For any matrix $\mathbf{A} \in \mathbb{C}^{M\times N}$, A_{ij} denotes the (i,j)th entry, while $\mathbf{A}^T$, and $\mathbf{A}^H$ return the transpose and the conjugate transpose of $\mathbf{A}$, respectively. For a square matrix $\mathbf{B}$, $\mathbf{B}^{\frac{1}{2}}$, $\mathbf{B}^{-1}$, $\mathsf{tr}(\mathbf{B})$, and $\det(\mathbf{B})$ denote the principal square root, inverse, trace, and determinant of $\mathbf{B}$, respectively. In addition, $\mathbf{I}_N$ is an $N\times N$ identity matrix, $\mathbf{0}_N$ denotes either an $N\times N$ zero matrix or a zero vector depending on the context, and $\mathbf{e}_i$ denotes the column vector with the ith element being 1 and 0 elsewhere. Finally, $\|\cdot\|_2$ represents the Euclidean norm of an input vector, and $\mathsf{E}\{\cdot\}$ returns the expectation of an input random entity.

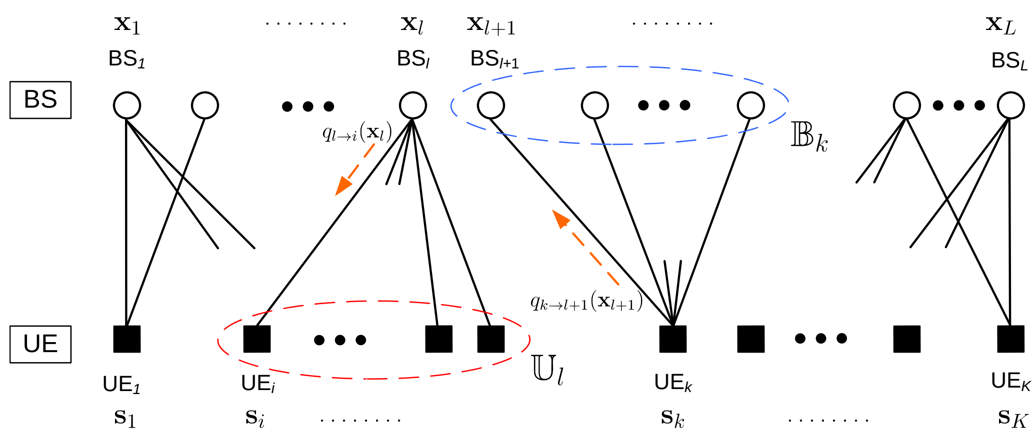

Figure 1: A downlink model with BS cooperation.

II. System Model and Problem Formulation

As shown in Figure 1, we consider a large-scale MIMO broadcast system where L interconnected multi-antenna BSs, labeled as $\mathsf{BS}_1, \ldots, \mathsf{BS}_L$, simultaneously send information to K users, labeled as $\mathsf{UE}_1, \ldots, \mathsf{UE}_K$. In the system, UE_k is equipped with M_k antennas while BS_l is equipped with N_l antennas. Let $M \triangleq \sum_{k=1}^{K} M_k$ and $N \triangleq \sum_{l=1}^{L} N_l$. The received signals at all the UEs can be expressed in a vector form as $\mathbf{y} = [\mathbf{y}_1^T, \ldots, \mathbf{y}_K^T]^T \in \mathbb{C}^N$, which is modeled as

$$\mathbf{y} = \begin{bmatrix} \mathbf{H}_{1,1} & \cdots & \mathbf{H}_{1,L} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{K,1} & \cdots & \mathbf{H}_{K,L} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_L \end{bmatrix} + \mathbf{z} \triangleq \mathbf{H}\mathbf{x} + \mathbf{z}, \tag{1}$$

where $\mathbf{x}_l$ denotes the transmitted signal from BS l, $\mathbf{H}_{k,l} \in \mathbb{C}^{M_k \times N_l}$ represents the channel matrix from BS l to UE k, and $\mathbf{z}$ is the complex Gaussian noise vector with zero mean and the covariance matrix $\sigma^2 \mathbf{I}_M$. In (1), we have defined $\mathbf{x} \in \mathbb{C}^N$ as the vector of the transmitted signal and $\mathbf{H} \in \mathbb{C}^{M \times N}$ as the overall downlink channel matrix. Although (1) appears to look like an $M \times N$ MIMO system, this is fundamentally different from a point-to-point MIMO channel. To see the differences, we emphasize the following two features.

First, note that $\mathbf{H}$ may have many zero block matrices because one UE is only able to receive signals from local BSs. The characteristic can be easily described via a graphical model as shown in Figure 1. For ease of expression, let $\mathbb{U}_l \subseteq \{1, 2, \ldots, K\}$ comprise the set of user indices such that BS_l has some inference on these UEs; i.e., $\mathbf{H}_{i,l} \neq \mathbf{0}$ for $i \in \mathbb{U}_l$. Similarly, let $\mathbb{B}_k \subseteq \{1, 2, \ldots, L\}$ be a set of BS indices such that these BSs have some inference on UE_k; i.e., $\mathbf{H}_{k,j} \neq \mathbf{0}$ for $j \in \mathbb{B}_k$. The local coupling is an important feature

that communication should only be required between a subset of BSs rather than among all BSs.

Secondly, since both BS and UE are equipped with multiple antennas, the spatial correlation of the MIMO channel for each link between a BS and a UE should be considered. In this paper, we employ the Kronecker model to characterize the spatial correlation of the MIMO channel for each link so that the correlation at a BS and a UE is modeled separately [35]. Specifically, the channel from BS_l to UE_k, $\mathbf{H}_{k,l} \in \mathbb{C}^{M_k \times N_l}$, can be written as

$$\mathbf{H}_{k,l} = \mathbf{R}_{k,l}^{\frac{1}{2}} \mathbf{W}_{k,l} \mathbf{T}_{k,l}^{\frac{1}{2}}, \tag{2}$$

where $\mathbf{R}_{k,l} \in \mathbb{C}^{M_k \times M_k}$ and $\mathbf{T}_{k,l} \in \mathbb{C}^{N_l \times N_l}$ are deterministic nonnegative definite matrices, which characterize the spatial correlation of the received signals across the antenna elements of UE_k and that of the transmitted signals across the antenna elements of BS_l respectively, and $\mathbf{W}_{k,l} \triangleq [\frac{1}{\sqrt{N_l}} W_{ij}^{(k,l)}] \in \mathbb{C}^{M_k \times N_l}$ consists of the random components of the channel in which the elements $\{W_{ij}^{(l,k)}\}_{1 \le i \le M_k; 1 \le j \le N_l}$ are i.i.d. complex Gaussian random variables with zero mean and unit variance. To get a proper definition on the channel gain of each link pair, we consider the power of the channel

$$\mathsf{E}\left\{\mathsf{tr}\left(\mathbf{H}_{k,l}\mathbf{H}_{k,l}^H\right)\right\} = \frac{1}{M_k}\mathsf{tr}\,(\mathbf{R}_k)\mathsf{tr}\,(\mathbf{T}_k). \tag{3}$$

If we assume that $\mathbf{R}_{k,l}$ and $\mathbf{T}_{k,l}$ are normalized such that $\mathsf{tr}(\mathbf{R}_{k,l}) = \varrho_{k,l} N_l$ and $\mathsf{tr}(\mathbf{T}_{k,l}) = M_k$, then $\varrho_{k,l}$ can be used as an indicator for the link gain between BS_l and UE_k.[1]

In the broadcast system (1), linear precoding, referred to as RZFBF, is used to project the data symbols onto a subspace using the N transmit antennas. Let $\mathbf{s} = [\mathbf{s}_1^T, \dots, \mathbf{s}_K^T]^T$ be the vector of data symbols, where $\mathbf{s}_k$ corresponds to the data symbols intended for UE_k. In RZFBF, the signal vector transmitted by the BSs, denoted by $\mathbf{x}$, is given by [16, 17]

$$\mathbf{x} = \alpha \mathbf{H}^H (\mathbf{H}\mathbf{H}^H + \beta \mathbf{I}_M)^{-1} \mathbf{s}, \tag{4}$$

where α is the normalization parameter to ensure that the transmit power constraint is met, i.e., $\mathsf{E}\{\|\mathbf{x}_l\|_2\} \le P_l$ for $l = 1, \dots, L$. Note that RZFBF can be regarded as a generalization of other beamformers by adjusting the regularization parameter β. For instance, if $\beta = 0$, it reduces to ZFBF whereas if $\beta \to \infty$, it will give the matched-filter beamforming. Several other beamformers can also have a RZFBF structure by designing the regularization parameter appropriately [18]. However, to obtain (4), all the BSs must cooperate to jointly process the data symbols from all the users in the network, requiring *global* CSI. If N and M are very large (which they should in order to benefit from the gains of MU-MIMO and BS cooperation),

[1] Indeed, the link gain can be included in either $\mathbf{R}_{k,l}$ or $\mathbf{T}_{k,l}$.

the centralized approach will become prohibitively complex. Therefore, in the next section, we propose message passing algorithms that can realize RZFBF in a distributed manner.

III. A Bayesian Approach to Distributed RZFBF

Now, we are concerning with the problem of distributing the computation of (4) among the BSs. Toward this end, we first use the *virtual* model concept of [21], which recasts the RZFBF optimization problem into an estimation problem which is described as follows.

Consider the virtual model

$$\mathbf{s} = \mathbf{H}\mathbf{x} + \tilde{\mathbf{z}}, \tag{5}$$

where $\tilde{\mathbf{z}} \in \mathbb{C}^M$ is the Gaussian random vector with zero mean and the covariance matrix $\beta \mathbf{I}_M$. Recall that $\mathbf{s}$ is the data symbol vector for all the users and $\mathbf{x}$ is the signal transmitted by the BSs. The virtual model implies that the transmitted signal $\mathbf{x}$ goes through the channel $\mathbf{H}$ and is then observed by $\mathbf{s} - \tilde{\mathbf{z}}$ at the UE sides. What is important here is that the virtual model allows us to process the beamforming problem through a probabilistic inference approach. Specifically, we adopt a Bayesian approach.

The Bayes optimal way of estimating $\mathbf{x}$ that minimizes the mean square error is given by [36]

$$\hat{\mathbf{x}} = \int \mathbf{x} p(\mathbf{x}|\mathbf{s}) d\mathbf{x}, \tag{6}$$

where $p(\mathbf{x}|\mathbf{s})$ is the posterior probability of $\mathbf{x}$ given observation of $\mathbf{s}$. Following Bayes theorem, we have

$$p(\mathbf{x}|\mathbf{s}) = \frac{p(\mathbf{s}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{s})}, \tag{7}$$

where the conditional distribution of $\mathbf{s}$ given $\mathbf{x}$ under (5) is given by

$$p(\mathbf{s}|\mathbf{x}) = \frac{1}{(\pi\beta)^N} e^{-\frac{1}{\beta}\sum_{k=1}^{K} \|\mathbf{s}_k - \sum_{l\in\mathbb{B}_k} \mathbf{H}_{k,l}\mathbf{x}_l\|_2^2}. \tag{8}$$

If we assume that $\mathbf{x}$ is taken from the standard complex Gaussian random vector and its density is given by $p(\mathbf{x}) = \frac{e^{-\|\mathbf{x}\|_2^2}}{\pi^N}$, then the posterior distribution $p(\mathbf{x}|\mathbf{s})$ admits an explicit expression as

$$p(\mathbf{x}|\mathbf{s}) = \frac{1}{\mathsf{Z}}\, e^{-\frac{1}{\beta}\sum_{k=1}^{K} \|\mathbf{s}_k - \sum_{l\in\mathbb{B}_k} \mathbf{H}_{k,l}\mathbf{x}_l\|_2^2 - \sum_{l=1}^{L}\|\mathbf{x}_l\|_2^2}, \tag{9}$$

Henceforth, we shall use Z to denote a *universal* normalization factor whose value may vary from one appearance to another. Plugging (9) into (6) and applying the Gaussian integral (Lemma 1 in Appendix

B), one can get that the solution of (6) is exactly identical to the form of (4) without the power normalization parameter. Next, we shall use an approach called BP (belief-propagation) for computing (6).

A. BP-RZFBF

We begin by applying the standard BP algorithm [37] to perform (6). As a matter of fact, BP can be regarded as a graphical method to estimate the marginal distributions of the distribution $p(\mathbf{x}|\mathbf{s})$ with respect to the variables $\mathbf{x}_l$. To this end, we reformulate the problem as a bipartite graph called the factor graph. The corresponding factor graph is depicted in Figure 1 where a circle represents a variable node associated with the transmit beamforming vector; i.e., $\mathbf{x}_l$ for BS_l, whereas a square indicates a factor node associated with the sub-constraint function; i.e., $\|\mathbf{s}_k - \sum_{l\in\mathbb{B}_k}\mathbf{H}_{k,l}\mathbf{x}_l\|_2^2$ for UE_k. There is an edge between a variable node l and a function node k if and only if $\mathbf{H}_{k,l} \neq \mathbf{0}$.

To estimate the marginal distributions, BP performs a set of message passing equations that go from factor nodes to variable nodes (i.e., $k \to l$) and from variable nodes to factor nodes (i.e., $l \to k$) as is illustrated in Figure 1. The message $q_{k\to l}$ from the factor node k to the variable node l is the marginal probability of the variable $\mathbf{x}_l$ when only the sub-constraint k is present. On the other hand, the message $q_{l\to k}$ from the variable node l to the factor node k is the marginal probability of the variables $\mathbf{x}_l$ in the absence of the sub-constraint k.

Specifically, in order to estimate these marginal distributions $p(\mathbf{x}_l|\mathbf{s})$ with BP algorithm, $2KL$ messages for the probability distributions of the variables $\mathbf{x}_l$ are constructed in the following way [37]:

$$q_{k\to l}^{(t)}(\mathbf{x}_l) = \frac{1}{Z_{k\to l}}\int \prod_{j\in\mathbb{B}_k\setminus l} d\mathbf{x}_j q_{j\to k}^{(t-1)}(\mathbf{x}_j) e^{-\frac{1}{\beta}\|\mathbf{s}_k - \sum_{j\in\mathbb{B}_k\setminus l}\mathbf{H}_{k,j}\mathbf{x}_j - \mathbf{H}_{k,l}\mathbf{x}_l\|_2^2}, \tag{10a}$$

$$q_{l\to k}^{(t-1)}(\mathbf{x}_l) = \frac{1}{Z_{l\to k}} p(\mathbf{x}_l) \prod_{i\in\mathbb{U}_l\setminus k} q_{i\to l}^{(t-1)}(\mathbf{x}_l), \tag{10b}$$

where $t = 1, 2, \ldots$ represents the iteration index and $Z^{k\to l}$ and $Z^{l\to k}$ are the normalization factors ensuring that $\int d\mathbf{x}_l q_{k\to l}^{(t)}(\mathbf{x}_l) = \int d\mathbf{x}_l q_{l\to k}^{(t-1)}(\mathbf{x}_l) = 1$. At the termination of the message passing algorithm, say at iteration T, the final estimate of $\mathbf{x}_l$ is given by $\hat{\mathbf{x}}_l = \int \mathbf{x}_l q_l^{(T)}(\mathbf{x}_l) d\mathbf{x}_l$ where $q_l^{(T)}(\mathbf{x}_l) \propto \prod_{k=1}^{K} q_{k\to l}^{(T)}(\mathbf{x}_l)$. The RZFBF solution can thus be realized in a distributed manner via the message passing procedures. However, the messages are *density functions* which are usually too complex to be exchanged and will cost a huge burden in the backhaul in our application of BS cooperation.

To overcome this, the message can be approximated by Gaussian and parameterized by the mean and

covariance. Instead of passing the density functions, we thus have the mean and covariance as the messages:

$$\mathbf{x}_{l\to k}^{(t)} = \langle \mathbf{x}_l \rangle_{q_{l\to k}^{(t)}}, \tag{11}$$

$$\mathbf{V}_{l\to k}^{(t)} = \left\langle (\mathbf{x}_l - \mathbf{x}_{l\to k}^{(t)})(\mathbf{x}_l - \mathbf{x}_{l\to k}^{(t)})^H \right\rangle_{q_{l\to k}^{(t)}}, \tag{12}$$

where $\langle f(\mathbf{x}_l)\rangle_{q_{l\to k}}$ denotes the average or expectation of a function $f(\mathbf{x}_l)$ over the random vector $\mathbf{x}_l$ with distribution $q_{l\to k}(\mathbf{x}_l)$. Mathematically, that is

$$\langle f(\mathbf{x}_l)\rangle_{q_{l\to k}} \triangleq \int f(\mathbf{x}_l) q_{l\to k}(\mathbf{x}_l) d\mathbf{x}_l.$$

The Gaussian approximation method was introduced in [30, 38] when the message is scalar and the concerned matrix is sparse. In our case, we follow the techniques in [29] by considering that the block matrix $\mathbf{H}_{k,l}$ scales as $O(1/\sqrt{N_l})$. As a consequence, we can approximate $q_{k\to l}(\mathbf{x}_l)^{(t)}$ by

$$\tilde{q}_{k\to l}^{(t)}(\mathbf{x}_l) \propto e^{-\left(\mathbf{x}_l^H \mathbf{E}_{k\to l}^{(t)} \mathbf{x}_l - (\mathbf{F}_{k\to l}^{(t)})^H \mathbf{x}_l - \mathbf{x}_l^H \mathbf{F}_{k\to l}^{(t)}\right)}, \tag{13}$$

where

$$\mathbf{E}_{k\to l}^{(t)} = \mathbf{H}_{k,l}^H \left(\sum_{j\in\mathbb{B}_k\setminus l} \mathbf{H}_{k,j} \mathbf{V}_{j\to k}^{(t-1)} \mathbf{H}_{k,j}^H + \beta \mathbf{I}_{M_k} \right)^{-1} \mathbf{H}_{k,l}, \tag{14}$$

$$\mathbf{F}_{k\to l}^{(t)} = \mathbf{H}_{k,l}^H \left(\sum_{j\in\mathbb{B}_k\setminus l} \mathbf{H}_{k,j} \mathbf{V}_{j\to k}^{(t-1)} \mathbf{H}_{k,j}^H + \beta \mathbf{I}_{M_k} \right)^{-1} \left(\mathbf{s}_k - \sum_{j\in\mathbb{B}_k\setminus l} \mathbf{H}_{k,j} \mathbf{x}_{j\to k}^{(t)} \right). \tag{15}$$

Notice that $\mathbf{x}_{j\to k}$ and $\mathbf{V}_{j\to k}$ in (14)–(15) are functions of $q_{j\to k}$ which is also altered due to the approximation $\tilde{q}_{k\to l}$. To make the connection, from (13) and (10b), we have

$$\tilde{q}_{l\to k}^{(t)}(\mathbf{x}_l) \propto p(\mathbf{x}_l) e^{-\sum_{i\in\mathbb{U}_l\setminus k}\left(\mathbf{x}_l^H \mathbf{E}_{i\to l}^{(t)} \mathbf{x}_l - (\mathbf{F}_{i\to l}^{(t)})^H \mathbf{x}_l - \mathbf{x}_l^H \mathbf{F}_{i\to l}^{(t)}\right)}. \tag{16}$$

Henceforth, we will replace $\mathbf{x}_{l\to k}^{(t)}$ and $\mathbf{V}_{l\to k}^{(t)}$ in (14)–(15) by $\tilde{\mathbf{x}}_{l\to k}^{(t)}$ and $\tilde{\mathbf{V}}_{l\to k}^{(t)}$ which are, respectively, the mean and covariance over the probability distribution $\tilde{q}_{l\to k}^{(t)}$. Also, in the sequel, we will no longer use the probability distribution $q_{l\to k}^{(t)}$. However, for notational convenience, we will abuse our notation slightly and still use $\mathbf{x}_{l\to k}^{(t)}$ and $\mathbf{V}_{l\to k}^{(t)}$ to denote those mean and covariance over the probability distribution $\tilde{q}_{l\to k}^{(t)}$.

Recall that $\mathbf{x}_l$ is taken from the standard complex Gaussian random vector. By applying the Gaussian integral (Lemma 1 in Appendix B), $\mathbf{x}_{l\to k}$ and $\mathbf{V}_{l\to k}$ with the distribution $\tilde{q}_{l\to k}(\mathbf{x}_l)$ in (16) can be computed

analytically. These lead to the following closed form of the BP update:

$$\mathbf{x}_{l\to k}^{(t)} = \left(\overline{\mathbf{E}}_{l\backslash k}^{(t)} + \mathbf{I}_{N_l}\right)^{-1} \overline{\mathbf{F}}_{l\backslash k}^{(t)}, \tag{17a}$$

$$\mathbf{V}_{l\to k}^{(t)} = \left(\overline{\mathbf{E}}_{l\backslash k}^{(t)} + \mathbf{I}_{N_l}\right)^{-1}, \tag{17b}$$

where $\overline{\mathbf{E}}_{l\backslash k}^{(t)} \triangleq \sum_{i\in\mathbb{U}_l\backslash k} \mathbf{E}_{i\to l}^{(t)}$ and $\overline{\mathbf{F}}_{l\backslash k}^{(t)} \triangleq \sum_{i\in\mathbb{U}_l\backslash k} \mathbf{F}_{i\to l}^{(t)}$. The number of messages is still $2KL$. However, the message update here is only on the mean and covariance rather than the functional update in (10). At the termination of the BP, the final estimation of $\mathbf{x}_l$ is given by

$$\mathbf{x}_{l}^{(t)} = \left(\overline{\mathbf{E}}_{l}^{(t)} + \mathbf{I}_{N_l}\right)^{-1} \overline{\mathbf{F}}_{l}^{(t)}, \tag{18}$$

where $\overline{\mathbf{E}}_{l}^{(t)} \triangleq \sum_{i\in\mathbb{U}_l} \mathbf{E}_{i\to l}^{(t)}$ and $\overline{\mathbf{F}}_{l}^{(t)} \triangleq \sum_{i\in\mathbb{U}_l} \mathbf{F}_{i\to l}^{(t)}$. We refer to this algorithm as BP-RZFBF although a variant of BP is adopted here. In some applications, the regularization parameter β varies for different UEs. In these cases, we only have to simply replace β with β_k in (14)–(15).

BP-RZFBF is a generalization of [21, (27)–(28)] in which $\mathbf{H}_{k,l}$'s are scalars (real numbers). Clearly, this generalization can be applied to a wide range of scenarios with complex-valued systems. Additionally, it performs block matrix computations resulting in a natural partition of BSs.

B. AMP-RZFBF

In BP-RZFBF, each BS has to send separate messages with respect to k; i.e., $\mathbf{x}_{l\to k}^{(t)}$ and $\mathbf{V}_{l\to k}^{(t)}$ $\forall k$. We can reduce the messaging overhead to $2(K+L)$. To do so, we note that the messages $\mathbf{x}_{l\to k}^{(t)}$ and $\mathbf{V}_{l\to k}^{(t)}$ are functions of $\overline{\mathbf{E}}_{l\backslash k}^{(t)}$ and $\overline{\mathbf{F}}_{l\backslash k}^{(t)}$ which are nearly independent of k. However, one must keep all the correction terms that are linear in $\mathbf{H}_{l,k}$. This methodology was first introduced in compressed sensing applications in [27] and is referred to as AMP. Using AMP in the BF-RZFBF problem, we have developed the AMP-RZFBF algorithm in Algorithm 1. For readability, we give the detailed derivation in Appendix A.

Now, we turn our attention to realizing AMP-RZFBF for the cooperative system. In general, each iteration requires a broadcast and gathering operation. We assume that each BS has local data information and CSI; e.g., only $\{\mathbf{s}_k, \mathbf{H}_{l,k}\}$ for $k \in \mathbb{U}_l$ are known at BS_l. The first two steps of AMP-RZFBF consist of performing $\boldsymbol{\Omega}_k^{(t)}$ and $\boldsymbol{\nu}_k^{(t)}$ updates at BS_l. Notice that for BS_l, $\boldsymbol{\Omega}_k^{(t)}$ and $\boldsymbol{\nu}_k^{(t)}$ updates are only for indices $k \in \mathbb{U}_l$ which correspond to the user indices within its reception range. In order to update $\boldsymbol{\Omega}_k^{(t)}$ and $\boldsymbol{\nu}_k^{(t)}$, BS_l must gather $\mathbf{H}_{k,l}\mathbf{V}_l^{(t-1)}\mathbf{H}_{k,l}^H$ and $\mathbf{H}_{k,l}\mathbf{x}_l^{(t-1)}$ from the set of its neighboring BSs $\mathbb{B}_k$. After getting $\boldsymbol{\Omega}_k^{(t)}$ and $\boldsymbol{\nu}_k^{(t)}$, BS_l is able to compute $(\boldsymbol{\Sigma}_l^{(t)}, \boldsymbol{\mu}_l^{(t)})$ and then update $(\mathbf{x}_l^{(t)}, \mathbf{V}_l^{(t)})$ subsequently. Once $(\mathbf{x}_l^{(t)}, \mathbf{V}_l^{(t)})$ are

Algorithm 1: AMP-RZFBF

Input: Data symbols $\mathbf{s}_k$ for $k=1,\ldots,K$, channel matrices $\mathbf{H}_{k,l}$ for $k=1,\ldots,K$ and $l=1,\ldots,L$.
Output: Return the RZFBF $\mathbf{x}_l$ for $l=1,\ldots,L$.

begin
  Select $\mathbf{x}_l^{(0)}=\mathbf{0}$, $\mathbf{V}_l^{(0)}=\mathbf{I}_{N_l}$, and $\boldsymbol{\nu}_k^{(0)}=\mathbf{s}_k$ for $k=1,\ldots,K$ and $l=1,\ldots,L$;
  $t \Longleftarrow 1$
  repeat
    $\boldsymbol{\Omega}_k^{(t)}=\sum_{l\in\mathbb{B}_k}\mathbf{H}_{k,l}\mathbf{V}_l^{(t-1)}\mathbf{H}_{k,l}^H$;
    $\boldsymbol{\nu}_k^{(t)}=\mathbf{s}_k-\sum_{l\in\mathbb{B}_k}\mathbf{H}_{k,l}\mathbf{x}_l^{(t-1)}+\left(\boldsymbol{\Omega}_k^{(t-1)}+\beta\mathbf{I}_{M_k}\right)^{-1}\boldsymbol{\Omega}_k^{(t)}\boldsymbol{\nu}_k^{(t-1)}$;
    $\boldsymbol{\Sigma}_l^{(t)}=\sum_{k\in\mathbb{U}_l}\mathbf{H}_{k,l}^H\left(\boldsymbol{\Omega}_k^{(t)}+\beta\mathbf{I}_{M_k}\right)^{-1}\mathbf{H}_{k,l}$;
    $\boldsymbol{\mu}_l^{(t)}=\mathbf{x}_l^{(t-1)}+\left(\boldsymbol{\Sigma}_l^{(t)}\right)^{-1}\left[\sum_{k\in\mathbb{U}_l}\mathbf{H}_{k,l}^H\left(\boldsymbol{\Omega}_k^{(t)}+\beta\mathbf{I}_{M_k}\right)^{-1}\boldsymbol{\nu}_k^{(t)}\right]$;
    $\mathbf{x}_l^{(t)}=\left(\boldsymbol{\Sigma}_l^{(t)}+\mathbf{I}_{N_l}\right)^{-1}\boldsymbol{\Sigma}_l^{(t)}\boldsymbol{\mu}_l^{(t)}$;
    $\mathbf{V}_l^{(t)}=\left(\boldsymbol{\Sigma}_l^{(t)}+\mathbf{I}_{N_l}\right)^{-1}$;
    $t \Longleftarrow t+1$
  until *Predefined number of iterations is met*;

computed, BS_l will broadcast $\mathbf{H}_{k,l}\mathbf{V}_l^{(t)}\mathbf{H}_{k,l}^H$ and $\mathbf{H}_{k,l}\mathbf{x}_l^{(t)}$ to its neighboring BSs. The algorithm continues to repeat the procedures above until it reaches a predefined number of iterations.

In AMP-RZFBF, the computation of $\boldsymbol{\Sigma}_l^{(t)}$ and $\mathbf{V}_k^{(t)}$ involves several matrix inversions for every channel realization. These demand high computational cost and rapid information exchange between the BSs. To remedy this, we propose to infer these parameters based on CCoI which varies much slower than CSI.

C. CCoI-aided AMP-RZFBF

Starting from the initial condition, we approximate $\boldsymbol{\Omega}_l^{(1)}$ by its average with respect to different realization of the measurement matrix $\mathbf{H}_{k,l}$:

$$\boldsymbol{\Omega}_k^{(1)} \approx \mathsf{E}\left\{\boldsymbol{\Omega}_k^{(1)}\right\} = \sum_{l\in\mathbb{B}_k}\frac{1}{N_l}\mathsf{tr}(\mathbf{T}_{k,l})\mathbf{R}_{k,l}, \tag{19}$$

where the equality follows from Lemma 2 in Appendix B. The approximation is benefited by the self-averaging property in statistical physics; that is, a quantity per degree of freedom has small deviations from its mean. In fact, using techniques from random matrix theory, e.g., [39], one can show that as $N_l\to\infty$, $\boldsymbol{\Omega}_k^{(1)}\to\mathsf{E}\{\boldsymbol{\Omega}_k^{(1)}\}$ almost surely. We find it useful to denote $\tilde{\varsigma}_{k,l}^{(1)}=\frac{1}{N_l}\mathsf{tr}(\mathbf{T}_{k,l})$ and define

$$\overline{\mathbf{R}}_k^{(t)} \triangleq \sum_{l\in\mathbb{B}_k}\tilde{\varsigma}_{k,l}^{(t)}\mathbf{R}_{k,l}. \tag{20}$$

Applying the similar argument to $\mathbf{\Sigma}_l^{(t)}$ in Line 7 of Algorithm 1 for $t = 1$, we have

$$\mathbf{\Sigma}_l^{(t)} \approx \sum_{k\in\mathbb{U}_l} \frac{1}{N_l} \mathsf{tr}\Big(\mathbf{R}_{k,l}\left(\overline{\mathbf{R}}_k^{(t)} + \beta \mathbf{I}_{M_k}\right)^{-1}\Big)\mathbf{T}_{k,l}. \tag{21}$$

Again, for ease of expression, we also define

$$\varsigma_{k,l}^{(t)} \triangleq \frac{1}{N_l}\mathsf{tr}\Big(\mathbf{R}_{k,l}(\overline{\mathbf{R}}_k^{(t)} + \beta\mathbf{I}_{M_k})^{-1}\Big), \tag{22}$$

and

$$\overline{\mathbf{T}}_l^{(t)} \triangleq \sum_{k\in\mathbb{U}_l} \varsigma_{k,l}^{(t)}\mathbf{T}_{k,l}. \tag{23}$$

Substituting the above definitions, (21) is then expressed as

$$\mathbf{\Sigma}_l^{(t)} \approx \overline{\mathbf{T}}_l^{(t)}. \tag{24}$$

Now, $\mathbf{x}_l^{(t)}$ and $\mathbf{V}_l^{(t)}$ can be calculated as those in Lines 9–10 of Algorithm 1 but $\mathbf{V}_l^{(t)} = \left(\mathbf{\Sigma}_l^{(t)} + \mathbf{I}_{N_l}\right)^{-1}$ is approximated by

$$\mathbf{V}_l^{(t)} \approx \left(\overline{\mathbf{T}}_l^{(t)} + \mathbf{I}_{N_l}\right)^{-1}. \tag{25}$$

Note that when $t = 1$, $\mathbf{\Omega}_l^{(t)}$ is given by (19). Let us go ahead on the next round of iteration to get a general expression for $\mathbf{\Omega}_l^{(t)}$ for general t. Following the similar argument as that used in (19), we have

$$\mathbf{\Omega}_k^{(t)} \approx \mathsf{E}\left\{\mathbf{\Omega}_k^{(t)}\right\} = \sum_{l\in\mathbb{B}_k} \frac{1}{N_l}\mathsf{tr}\left(\mathbf{T}_{k,l}\left(\overline{\mathbf{T}}_k^{(t-1)} + \mathbf{I}_{N_l}\right)^{-1}\right)\mathbf{R}_{k,l}. \tag{26}$$

Define

$$\tilde{\varsigma}_{k,l}^{(t)} \triangleq \frac{1}{N_l}\mathsf{tr}\Big(\mathbf{T}_{k,l}\left(\overline{\mathbf{T}}_l^{(t-1)} + \mathbf{I}_{N_l}\right)^{-1}\Big). \tag{27}$$

Then (26) becomes

$$\mathbf{\Omega}_l^{(t)} \approx \sum_{k\in\mathbb{U}_l} \tilde{\varsigma}_{k,l}^{(t)}\mathbf{R}_{k,l}. \tag{28}$$

Recall that the updates of $\mathbf{\Omega}_l^{(t)}$, $\mathbf{\Sigma}_l^{(t)}$, and $\mathbf{V}_l^{(t)}$ in Lines 5, 7, and 10 of Algorithm 1, respectively, involve the channel realizations $\{\mathbf{H}_{k,l}\}$. These computations are replaced by (28), (24), and (25), where only the CCoI is required. Therefore, Algorithm 1 together with these replacements lead to the simpler iteration forms. The algorithmic description of this CCoI-aided AMP-RZFBF is summarized in Algorithm 2.

Algorithm 2: CCI-aided AMP-RZFBF

Input: Data symbols $\mathbf{s}_k$ for $k=1,\ldots,K$, channel matrices $\mathbf{H}_{k,l}$ for $k=1,\ldots,K$ and $l=1,\ldots,L$, and CCI $\{\mathbf{T}_{k,l},\mathbf{R}_{k,l}\}$ for $k=1,\ldots,K$ and $l=1,\ldots,L$.

Output: Return the RZFBF $\mathbf{x}_l$ for $l=1,\ldots,L$

1 **begin**

2 Select $\mathbf{x}_l^{(0)}=\mathbf{0}$, $\boldsymbol{\nu}_k^{(0)}=\mathbf{s}_k$, $\overline{\mathbf{R}}_k^{(0)}=\mathbf{0}$, and $\overline{\mathbf{T}}_l^{(0)}=\mathbf{0}$ for $k=1,\ldots,K$ and $l=1,\ldots,L$;

3 $t \Longleftarrow 1$

4 **repeat**

5 $\tilde{\varsigma}_{k,l}^{(t)} = \frac{1}{N_l}\mathsf{tr}\left(\mathbf{T}_{k,l}\left(\overline{\mathbf{T}}_l^{(t-1)}+\mathbf{I}_{N_l}\right)^{-1}\right)$;

6 $\overline{\mathbf{R}}_k^{(t)} = \sum_{l\in\mathbb{B}_k}\tilde{\varsigma}_{k,l}^{(t)}\mathbf{R}_{k,l}$;

7 $\varsigma_{k,l}^{(t)} = \frac{1}{N_l}\mathsf{tr}\left(\mathbf{R}_{k,l}\left(\overline{\mathbf{R}}_k^{(t)}+\beta\mathbf{I}_{M_k}\right)^{-1}\right)$;

8 $\overline{\mathbf{T}}_l^{(t)} = \sum_{k\in\mathbb{U}_l}\varsigma_{k,l}^{(t)}\mathbf{T}_{k,l}$;

9 $\mathbf{A}_l^{(t)} = \overline{\mathbf{T}}_l^{(t)}\left(\overline{\mathbf{T}}_l^{(t)}+\mathbf{I}_{N_l}\right)^{-1}$;

10 $\mathbf{B}_k^{(t)} = \overline{\mathbf{R}}_k^{(t)}\left(\overline{\mathbf{R}}_k^{(t-1)}+\beta\mathbf{I}_{M_k}\right)^{-1}$;

11 $\boldsymbol{\nu}_k^{(t)} = \mathbf{s}_k - \sum_{l\in\mathbb{B}_k}\mathbf{H}_{k,l}\mathbf{x}_l^{(t-1)} + \mathbf{B}_k^{(t)}\boldsymbol{\nu}_k^{(t-1)}$;

12 $\mathbf{x}_l^{(t)} = \mathbf{A}_l^{(t)}\left[\mathbf{x}_l^{(t-1)} + \left(\overline{\mathbf{T}}_l^{(t)}\right)^{-1}\left(\sum_{k\in\mathbb{U}_l}\mathbf{H}_{k,l}^H\left(\overline{\mathbf{R}}_k^{(t)}+\beta\mathbf{I}_{M_k}\right)^{-1}\boldsymbol{\nu}_k^{(t)}\right)\right]$;

13 $t \Longleftarrow t+1$

14 **until** *Predefined number of iterations is met*;

The realization of CCoI-aided AMP-RZFBF is similar to that of AMP-RZFBF but with much lower computational complexity and much less communication overhead. Firstly, notice that lines 5–10 of Algorithm 2 can be computed *offline* and *locally* regardless of the channel realizations $\{\mathbf{H}_{k,l}\}$, data symbols $\{\mathbf{s}_k\}$, and the outputs $\{\mathbf{x}_l^{(t)}\}$ of each iteration. Because CCoI can be considered static, the BSs compute and exchange these parameters at the time scale at which the CCoI changes rather than the instantaneous channel realizations. This characteristics significantly reduces the computational complexity and the communication overhead. Secondly, the remaining two steps, lines 11–12 of Algorithm 2, involve only linear matrix multiplications. The update of $\boldsymbol{\nu}_k^{(t)}$ and $\mathbf{x}_l^{(t)}$ also requires a general broadcast and gathering operation. In particular, to update $\boldsymbol{\nu}_k^{(t)}$, BS_l must gather $\mathbf{H}_{k,l}\mathbf{x}_l^{(t-1)}$ from the set of BSs $\mathbb{B}_k$ but it only updates $\boldsymbol{\nu}_k^{(t)}$ for $k\in\mathbb{U}_l$. After $\mathbf{x}_l^{(t)}$ is computed, BS_l will broadcast $\mathbf{H}_{k,l}\mathbf{x}_l^{(t)}$ to its neighboring BSs. The algorithm continues to repeat the procedures above until it reaches a predefined number of iterations.

IV. Simulation Results

In this section, we compare the performance of different algorithms through simulations. The considered algorithms include all the message passing algorithms in Section III (i.e., BP-RZFBF, AMP-RZFBF, and CCoI-aided AMP-RZFBF) and the ADMM approach in [33, Section 8.3]. ADMM is the state-of-the-art

optimization technique and has now been widely used in performing distributed estimations.

Before proceeding, let us first take a look at the computational complexity of these algorithms. In BP-RZFBF, most of computational complexity lies in the matrix inversions in (14)–(15) and (17). Moreover, we have to perform these matrix inversions for all the $2KL$ messages. This gives a complexity of order $KLO(M_k^3)$ for each iteration. In AMP-RZFBF, the complexity also lies in the matrix inversions while the messaging overhead is reduced to $2(K+L)$. Therefore, the complexity of AMP-RZFBF is of order $(K+L)O(M_k^3)$ for each iteration. The complexity of the ADMM approach is comparable to AMP-RZFBF. Finally, the complexity of CCoI-aided AMP-RZFBF is further reduced from AMP-RZFBF because the matrix inversions are performed at the time scale at which the CCoI changes. Therefore, the computational complexity of CCoI-aided AMP-RZFBF is of order $\left(\frac{K+L}{\tau}\right) O(M_k^3)$ for each iteration where τ represents the time scale at which the CCoI changes. The value of τ could be very large because CCoI can be considered static. Consequently, CCoI-aided AMP-RZFBF can be implemented in the most efficient way.

With the computational complexity in mind, our attention turns to their performances. We consider a cellular system with 100 BSs and 100 UEs in which each BS is equipped with 8 transmit antennas and each user has 4 receive antennas, i.e., $L = 100$, $K = 100$, $N_l = 8$, and $M_k = 4$. The propagation channel matrix between each BS and UE is characterized by (2), where the spatial correlations $\mathbf{R}_{k,l}$'s and $\mathbf{T}_{k,l}$'s are arbitrarily generated with elements being $[\mathbf{R}_{k,l}]_{i,j} = \rho_{\mathsf{R}_{k,l}}^{|i-j|}$ and $\mathbf{T}_{k,l}]_{i,j} = \rho_{\mathsf{T}_{k,l}}^{|i-j|}$, respectively. Additionally, the link gain $\varrho_{k,l}$ is included in $\mathbf{R}_{k,l}$ and is also uniformly and randomly generated. Figure 2 illustrates the average throughput of the algorithms varies with the number of message transfers. The average throughput is calculated by $\frac{1}{M}\sum_{k=1}^{K}\sum_{m=1}^{M_k}\log_2\left(1+\gamma_{m,k}^{(t)}\right)$ where $\gamma_{m,k}^{(t)} \triangleq \frac{|\mathbf{e}_m^T\mathbf{s}_k|^2}{|\mathbf{e}_m^T(\mathbf{H}_k\mathbf{x}^{(t)}-\mathbf{s}_k)|^2+\sigma^2}$ and $\mathbf{x}^{(t)}$ is the vector of transmitted signals at the t-th iteration. Here, $\mathbf{H}_k$ denotes $[\mathbf{H}_{k,1}\cdots\mathbf{H}_{k,L}]$ and $\mathbf{e}_m$ has been defined in Notations. The results provided are for a particular realization of the channel. It is natural that when the number of iterations increases, the average throughput increases and saturates eventually. Here, RZFBF in (4) serves as a benchmark for the optimal beamformer. From Figure 2, it can be observed that the proposed message passing algorithms converge significantly faster than the ADMM approach. The convergence rates of all the proposed message passing algorithms are very similar.

Recall that AMP-RZFBF follows from BP-RZFBF but using the approximations that $\overline{\mathbf{E}}_{l\setminus k}$ and $\overline{\mathbf{F}}_{l\setminus k}$ are nearly independent of k. This approximation is expected to be good if K and L are extremely large. Furthermore, the CCoI-aided AMP-RZFBF uses the large system approximation by assuming $N_l \to \infty$. Although the setting in Figure 2 corresponds to a practical system dimension, it is intriguing to see their performances under a relatively small network; e.g., $L = 16$, $K = 16$, $N_l = 4$, and $M_k = 2$. Under the small network consideration, Figure 3 illustrates the convergence of the algorithms. Similar characteristics

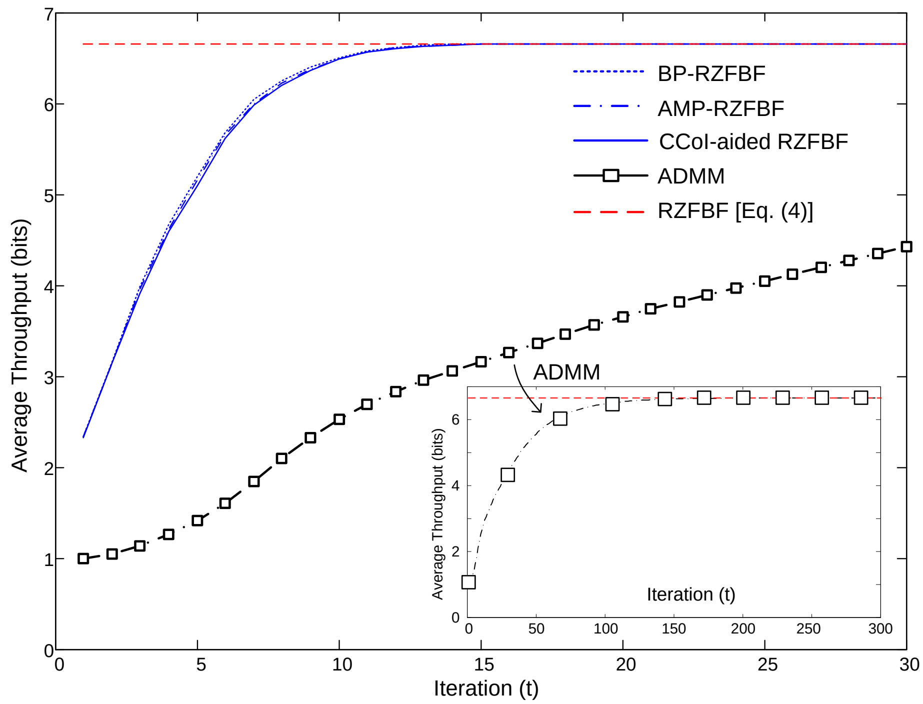

Figure 2: Average throughput against the number of iterations for the message passing beamformers and global beamformer when $L = 100$, $K = 100$, $N_l = 8$, $M_k = 4$, and $\beta = \sigma^2 = 10^{-2}$.

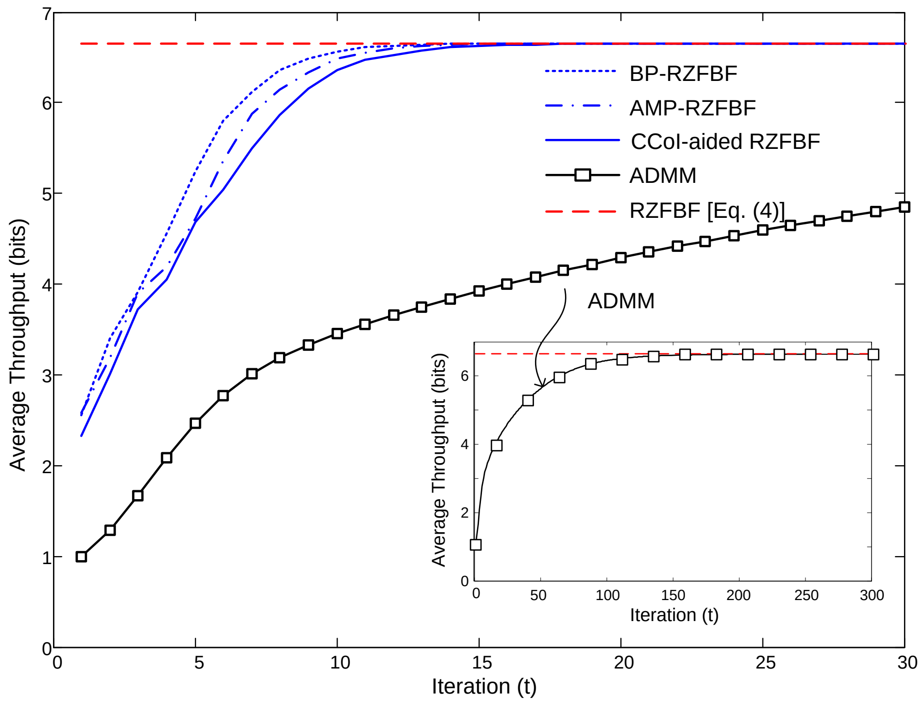

Figure 3: Average throughput against the number of iterations for the message passing beamformers and global beamformer when $L = 16$, $K = 16$, $N_l = 4$, $M_k = 2$, and $\beta = \sigma^2 = 10^{-2}$.

as in Figure 2 before are observed. Additionally, comparing to BP-RZFBF, AMP-RZFBF and CCoI-aided AMP-RZFBF only slightly degrades the convergence rate. This result is quite different from several earlier designs based on CCoI, e.g., [31, 32]. Usually, when some calculations are approximated by the CCoI, an obvious degradation in performance would be observed but this is not the case in our scheme.

V. Conclusion

Using Bayesian inference, this paper proposed several message passing algorithms for realizing RZFBF in cooperative-BS networks, namely, BP-RZFBF, AMP-RZFBF and CCoI-aided AMP-RZFBF. Results showed that the proposed algorithms converge very fast to the exact RZFBF. Comparing to BP-RZFBF, both AMP-RZFBF and CCoI-aided AMP-RZFBF perform well with only very slight degradation in the convergence rate, but greatly reducing the burden for information exchange between the BSs.

Appendix A: Derivation for AMP-RZFBF

To derive AMP-RZFBF, we use a heuristic approximation which keeps all the terms that are linear in the matrix $\mathbf{H}_{l,k}$ while neglecting the higher-order terms. The similar methodology was used in [29] in the case of compressed sensing although some modifications are required to reflect the concerned case.

We start by noticing that $\overline{\mathbf{E}}_{l\backslash k}^{(t)} = \sum_{i\neq k} \mathbf{E}_{i\to l}^{(t)}$ is the sum of K terms each of order $1/N_l$ because $\mathbf{H}_{l,k}$ scales as $O(1/\sqrt{N_l})$. Therefore, it is natural to approximate $\overline{\mathbf{E}}_{l\backslash k}^{(t)}$ by $\overline{\mathbf{E}}_{l}^{(t)} = \sum_{i} \mathbf{E}_{i\to l}^{(t)}$ which only depends on the index l and not on k. Similarly, it is natural to anticipate a similar approximation for $\overline{\mathbf{F}}_{l\backslash k}^{(t)}$. However, we must be careful to keep all correction terms of order $1/\sqrt{N_l}$. To that end, we instead set

$$\overline{\mathbf{F}}_{l\backslash k}^{(t)} = \overline{\mathbf{F}}_{l}^{(t)} - \boldsymbol{\Delta}\overline{\mathbf{F}}_{l\to k}^{(t)}. \tag{29}$$

Recall from (17) that $\mathbf{x}_{l\to k}^{(t)} = \left(\overline{\mathbf{E}}_{l\backslash k}^{(t)} + \mathbf{I}_{N_l}\right)^{-1} \overline{\mathbf{F}}_{l\backslash k}^{(t)}$. Then we get

$$\begin{aligned}\mathbf{x}_{l\to k}^{(t)} &\approx \left(\overline{\mathbf{E}}_{l}^{(t)} + \mathbf{I}_{N_l}\right)^{-1} \overline{\mathbf{F}}_{l}^{(t)} - \left(\overline{\mathbf{E}}_{l}^{(t)} + \mathbf{I}_{N_l}\right)^{-1} \boldsymbol{\Delta}\overline{\mathbf{F}}_{l\to k}^{(t)} \\ &= \mathbf{x}_{l}^{(t)} - \left(\overline{\mathbf{E}}_{l}^{(t)} + \mathbf{I}_{N_l}\right)^{-1} \boldsymbol{\Delta}\overline{\mathbf{F}}_{l\to k}^{(t)}. \end{aligned} \tag{30}$$

We will approximate the above two terms by dropping their negligible components. Before proceeding, we deal with the approximation of $\overline{\mathbf{E}}_{l}^{(t)}$. Let us define $\boldsymbol{\Omega}_{k}^{(t)} \triangleq \sum_j \mathbf{H}_{k,j}\mathbf{V}_{q_{j\to k}}^{(t-1)}\mathbf{H}_{k,j}^{H}$. Then we have

$$\begin{aligned}\overline{\mathbf{E}}_{l}^{(t)} &= \sum_{k} \mathbf{H}_{k,l}^{H} \left(\boldsymbol{\Omega}_{k}^{(t)} - \mathbf{H}_{k,l}\mathbf{V}_{l\to k}^{(t-1)}\mathbf{H}_{k,l}^{H} + \beta\mathbf{I}_{M_k}\right)^{-1} \mathbf{H}_{k,l} \\ &\approx \sum_{k} \mathbf{H}_{k,l}^{H} \left(\boldsymbol{\Omega}_{k}^{(t)} + \beta\mathbf{I}_{M_k}\right)^{-1} \mathbf{H}_{k,l} \triangleq \boldsymbol{\Sigma}_{l}^{(t)}, \end{aligned} \tag{31}$$

where the approximation follows from the fact that $\mathbf{H}_{k,l}\mathbf{V}_{l\to k}^{(t-1)}\mathbf{H}_{k,l}^{H}$ is of order $1/N_l$ and can be safely neglected. Similarly, we note that $\mathbf{V}_{l\to k}^{(t-1)}$ is nearly independent of k. This leads to

$$\mathbf{V}_{l\to k}^{(t-1)} = \left(\overline{\mathbf{E}}_{l\backslash k}^{(t-1)} + \mathbf{I}_{N_l}\right)^{-1} \approx \left(\boldsymbol{\Sigma}_{l}^{(t-1)} + \mathbf{I}_{N_l}\right)^{-1} \triangleq \mathbf{V}_{l}^{(t-1)}. \tag{32}$$

Then we get

$$\boldsymbol{\Omega}_{k}^{(t)} = \sum_{j} \mathbf{H}_{k,j}\mathbf{V}_{q_{j\to k}}^{(t-1)}\mathbf{H}_{k,j}^{H} \approx \sum_{j} \mathbf{H}_{k,j}\mathbf{V}_{j}^{(t-1)}\mathbf{H}_{k,j}^{H}. \tag{33}$$

Now, we return to the approximation of $\mathbf{x}_{l\to k}^{(t)}$. First, we deal with the second terms of (30) and get

$$\begin{aligned}
\mathbf{x}_{l\to k}^{(t)} &\approx \mathbf{x}_l^{(t)} - \left(\overline{\mathbf{E}}_l^{(t)} + \mathbf{I}_{N_l}\right)^{-1} \left[\mathbf{H}_{k,l}^H \left(\boldsymbol{\Omega}_k^{(t)} - \mathbf{H}_{k,l}\mathbf{V}_{l\to k}^{(t-1)}\mathbf{H}_{k,l}^H + \beta\mathbf{I}_{M_k}\right)^{-1} \left(\boldsymbol{\nu}_k^{(t)} + \mathbf{H}_{k,l}\mathbf{x}_{l\to k}^{(t-1)}\right)\right] \\
&\approx \mathbf{x}_l^{(t)} - \left(\boldsymbol{\Sigma}_l^{(t)} + \mathbf{I}_{N_l}\right)^{-1} \mathbf{H}_{k,l}^H \left(\boldsymbol{\Omega}_k^{(t)} + \beta\mathbf{I}_{M_k}\right)^{-1} \boldsymbol{\nu}_k^{(t)},
\end{aligned}$$

where the first approximation is directly from (30) by substituting the definition of $\boldsymbol{\Delta}\overline{\mathbf{F}}_{l\to k}^{(t)}$ and we have defined $\boldsymbol{\nu}_k^{(t)} \triangleq \mathbf{s}_k^{(t-1)} - \sum_j \mathbf{H}_{k,j}\mathbf{x}_{q_{j\to k}}^{(t-1)}$. Substituting the above approximation of $\mathbf{x}_{l\to k}^{(t)}$ in $\boldsymbol{\nu}_k^{(t)}$, we get

$$\begin{aligned}
\boldsymbol{\nu}_k^{(t)} &\approx \mathbf{s}_k^{(t-1)} - \sum_j \mathbf{H}_{k,j}\mathbf{x}_j^{(t-1)} - \left(\sum_j \mathbf{H}_{k,j}(\boldsymbol{\Sigma}_j^{(t-1)} + \mathbf{I}_{N_j})^{-1}\mathbf{H}_{k,j}^H\right) \left(\boldsymbol{\Omega}_k^{(t-1)} + \beta\mathbf{I}_{M_k}\right)^{-1} \boldsymbol{\nu}_k^{(t-1)} \\
&= \mathbf{s}_k^{(t-1)} - \sum_j \mathbf{H}_{k,j}\mathbf{x}_j^{(t-1)} - \boldsymbol{\Omega}_k^{(t)} \left(\boldsymbol{\Omega}_k^{(t-1)} + \beta\mathbf{I}_{M_k}\right)^{-1} \boldsymbol{\nu}_k^{(t-1)}, \qquad (34)
\end{aligned}$$

where the second equality follows from (33).

Now, it remains to complete the calculation of $\mathbf{x}_l^{(t)}$. We start from the definition

$$\mathbf{x}_l^{(t)} = \left[\left(\overline{\mathbf{E}}_l^{(t)}\right)^{-1} + \mathbf{I}_{N_l}\right]^{-1} \left(\overline{\mathbf{E}}_l^{(t)}\right)^{-1} \overline{\mathbf{F}}_l^{(t)} = \left[\left(\overline{\mathbf{E}}_l^{(t)}\right)^{-1} + \mathbf{I}_{N_l}\right]^{-1} \boldsymbol{\mu}_l^{(t)}, \qquad (35)$$

where we have defined

$$\boldsymbol{\mu}_l^{(t)} \triangleq \left(\overline{\mathbf{E}}_l^{(t)}\right)^{-1} \overline{\mathbf{F}}_l^{(t)}. \qquad (36)$$

Following the similar approximations as above, we get

$$\begin{aligned}
\boldsymbol{\mu}_l^{(t)} = &\left[\sum_k \mathbf{H}_{k,l}^H \left(\boldsymbol{\Omega}_k^{(t)} - \mathbf{H}_{k,l}\mathbf{V}_{l\to k}^{(t-1)}\mathbf{H}_{k,l}^H + \beta\mathbf{I}_{M_k}\right)^{-1} \mathbf{H}_{k,l}\right]^{-1} \\
&\times \left[\sum_k \mathbf{H}_{k,l}^H \left(\boldsymbol{\Omega}_k^{(t)} - \mathbf{H}_{k,l}\mathbf{V}_{l\to k}^{(t-1)}\mathbf{H}_{k,l}^H + \beta\mathbf{I}_{M_k}\right)^{-1} \left(\boldsymbol{\nu}_k^{(t)} + \mathbf{H}_{k,l}\mathbf{x}_{l\to k}^{(t-1)}\right)\right] \\
&\approx \mathbf{x}_l^{(t-1)} + \left(\boldsymbol{\Sigma}_l^{(t)}\right)^{-1} \left[\sum_k \mathbf{H}_{k,l}^H \left(\boldsymbol{\Omega}_k^{(t)} + \beta\mathbf{I}_{M_k}\right)^{-1} \boldsymbol{\nu}_k^{(t)}\right] \qquad (37)
\end{aligned}$$

and then

$$\mathbf{x}_l^{(t)} \approx \left(\boldsymbol{\Sigma}_l^{(t)} + \mathbf{I}_{N_l}\right)^{-1} \boldsymbol{\Sigma}_l^{(t)}\boldsymbol{\mu}_l^{(t)}. \qquad (38)$$

Putting the above relations (31), (33), (34), (37) and (38) together, we get AMP-RZFBF.

Appendix B: Lemmas

For convenience, we provide some mathematical tools needed in this paper.

Lemma 1 *Given a positive definite matrix* $\mathbf{A}$*, we have*

$$\frac{1}{\mathsf{Z}} \int \mathbf{x} e^{-\mathbf{x}^H \mathbf{A} \mathbf{x} + \mathbf{b}^H \mathbf{x} + \mathbf{x}^H \mathbf{b}} d\mathbf{x} = \mathbf{A}^{-1} \mathbf{b}, \tag{39}$$

where Z *is a normalization factor such that* $1/\mathsf{Z} \int e^{-\mathbf{x}^H \mathbf{A} \mathbf{x} + \mathbf{b}^H \mathbf{x} + \mathbf{x}^H \mathbf{b}} d\mathbf{x} = 1$.

Lemma 2 *A random matrix* $\mathbf{X} \in \mathbb{C}^{M \times N}$ *is said to have a matrix variate complex Gaussian distribution with mean* $\bar{\mathbf{X}}$ *and covariance matrix* $\mathbf{B} \otimes \mathbf{A}$*, if it can be written by* $\bar{\mathbf{X}} + \mathbf{A}^{\frac{1}{2}} \mathbf{W} \mathbf{B}^{\frac{1}{2}}$*, where* $\mathbf{A} \in \mathbb{C}^{N \times N}$ *and* $\mathbf{B} \in \mathbb{C}^{M \times M}$ *are both positive definite and the elements of* $\mathbf{W}$ *are i.i.d. complex Gaussian random variables with zero mean and unit variance. Then we have*

$$\mathsf{E}\{\mathbf{X}\mathbf{C}\mathbf{X}^H\} = \bar{\mathbf{X}}\mathbf{C}\bar{\mathbf{X}}^H + \mathsf{tr}(\mathbf{B}\mathbf{C})\mathbf{A}, \tag{40}$$

$$\mathsf{E}\{\mathbf{X}^H\mathbf{D}\mathbf{X}\} = \bar{\mathbf{X}}^H\mathbf{D}\bar{\mathbf{X}} + \mathsf{tr}(\mathbf{A}\mathbf{D})\mathbf{B}. \tag{41}$$

References

[1] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt, “An introduction to the multiuser MIMO downlink,” *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 60–67, Oct. 2004.

[2] G. Z. Pan, K. K. Wong, and T. S. Ng, “Generalized multiuser orthogonal space division multiplexing,” *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 1–5, Nov. 2004.

[3] D. Gesbert, M. Kountouris, R. W. Heath Jr., C.-B. Chae, and T. Salzer, “Shifting the MIMO paradigm,” *IEEE Sig. Proc. Mag.*, vol. 24, no. 5, pp. 36–46, Sep. 2007.

[4] S. Shamai (Shitz) and A. D. Wyner, “Information-theoretic considerations for symmetric, cellular, multiple-access fading channels – Part I & II,” *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1877–1911, Nov. 1997.

[5] G. Caire and S. Shamai (Shitz), “On the achievable throughput of a multiantenna Gaussian broadcast channel,” *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[6] O. Somekh, B. M. Zaidel, and S. Shamai (Shitz), “Sum rate characterization of joint multiple cell-site processing,” *IEEE Trans. Info. Theory*, vol. 53, no. 12, pp. 4473–4497, Dec. 2007.

[7] O. Somekh, O. Simeone, Y. Bar-Ness, A. M. Haimovich, and S. Shamai (Shitz), “Cooperative multicell zero-forcing beamforming in cellular downlink channels,” *IEEE Trans. Info. Theory*, vol. 2009, no. 7, pp. 3206–3219, Jul. 2009.

[8] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, “Multi-cell MIMO cooperative networks: A new look at interference,” *IEEE J. Sel. Areas Commun. Special Issue Cooperative Commun. MIMO Cellular Net.*, vol. 28, no. 9, pp. 1–29, Dec. 2010.

[9] C. K. Wen and K. K. Wong, "On the sum-rate of uplink MIMO cellular systems with amplify-and-forward relaying and collaborative base stations," *IEEE J. Sel. Areas Commun. Special Issue Cooperative Commun. MIMO Cellular Net.*, vol. 28, no. 9, pp. 1409–1424, Dec. 2010.

[10] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[11] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[12] D. Samardzija, H. Huang, T. Sizer, and R. Valenzuela, "Experimental downlink multiuser MIMO system with distributed and coherentlycoordinated transmit antennas," in *Proc. Int. Conference on Communications (ICC'09)*, Jun. 2007.

[13] R. Irmer, H.-P. Mayer, A. Weber, V. Braun, M. Schmidt, M. Ohm, N. Ahr, A. Zoch, C. Jandura, P. Marsch, and G. Fettweis, "Multisite field trial for LTE and advanced concepts," *IEEE Commun. Mag.*, vol. 47, no. 2, pp. 92–98, Feb. 2009.

[14] V. Jungnickel, A. Forck, S. Jaeckel, F. Bauermeister, S. Schiffermueller, S. Schubert, S. Wahls, L. Thiele, T. Haustein, W. Kreher, J. Mueller, H. Droste, and G. Kadel, "Field trials using coordinated multi-point transmission in the downlink," in *Proc. 3rd Int. Workshop on Wireless Distributed Networks (WDN)*, 2010.

[15] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[16] M. Joham, K. Kusume, M. H. Gzara, W. Utschick, and J. A. Nossek, "Transmit Wiener filter for the downlink of TDDDS-CDMA systems," in *Proc. IEEE 7th Int. Symp. Spread-Spectrum Tech. Appl. (ISSSTA)*, vol. 1, 2002, pp. 9–13.

[17] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication–Part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[18] R. Zakhour and S. Hanly, "Base station cooperation on the downlink: Large system analysis," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2079–2106, Apr. 2012.

[19] O. Simeone, N. Levy, A. Sanderovich, O. Somekh, B. M. Zaidel, H. V. Poor, and S. Shamai (Shitz), *Cooperative wireless cellular systems: An information-theoretic view.* Foundations and Trends in Commun. Inf. Theory, 2011.

[20] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[21] B. L. Ng, J. S. Evans, S. V. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5491–5499, Dec. 2008.

[22] E. Aktas, D. Aktas, S. V. Hanly and J. S. Evans, *Turbo Base Stations.* in Cooperative Cellular Wireless Networks (E. Hossain, D. In Kim and V. K. Bhargava, eds.), Cambridge University Press, 2011.

[23] E. Björnson, R. Zakhour, D. Gesbert, and B. Ottersten, "Cooperative multicell precoding: Rate region characterization and distributed strategies with instantaneous and statistical CSI," *IEEE Trans. Sig. Proc.*, vol. 58, no. 8, pp. 4298–4310, Aug. 2010.

[24] I. Sohn, S. H. Lee, and J. G. Andrews, "Belief propagation for distributed downlink beamforming in cooperative MIMO cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, pp. 4140–4149, Dec. 2011.

[25] Y. Huang, G. Zheng, M. Bengtsson, K.-K. Wong, L. Yang, and B. Ottersten, "Distributed multicell beamforming with limited intercell coordination," *IEEE Trans. Sig. Proc.*, vol. 59, no. 2, pp. 728–738, Feb. 2011.

[26] S. Rangan and R. Madan, "Belief propagation methods for intercell interference coordination in femtocell networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 631–640, Apr. 2012.

[27] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, 2009.

[28] M. Bayati and A. Montanari, “The dynamics of message passing on dense graphs, with applications to compressed sensing,,” *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.

[29] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, “Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices,” *J. Stat. Mech.*, vol. P08009, 2012.

[30] S. Rangan, “Generalized approximate message passing for estimation with random linear mixing,” *ArXiv e-prints:1010.5141v2 [cs.IT]*, 2010.

[31] J. Hoydis, M. Debbah, and M. Kobayashi, “Asymptotic moments for interference mitigation in correlated fading channels,” in *IEEE Internatinal Sym. Inf. Theory (ISIT'11)*, Saint-Petersburg, Russia, 2011.

[32] S. Lakshminarayana, M. Debbah, and M. Assaad, “Asymptotic analysis of distributed multi-cell beamforming,” *IEEE Trans. Sig. Proc.*, 2012, submitted.

[33] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers.* Foundations and Trends in Machine Learning, 2011.

[34] H. Zhu, A. Cano, and G. Giannakis, “Distributed consensus-based demodulation: algorithms and error analysis,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 6, pp. 2044–2054, June 2010.

[35] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, “Fading correlation and its effect on the capacity of multi-element antenna systems,” *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[36] H. V. Poor, *An Introduction to Signal Detection and Estimation.* New York: Springer-Verlag, 1994.

[37] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, “Factor graphs and the sum-product algorithm,” *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[38] D. Guo and C.-C. Wang, “Asymptotic mean-square optimality of belief propagation for sparse linear systems,” in *in Proc. IEEE Inform. Theory Workshop*, Chengdu, China, Oct. 2006, pp. 194–198.

[39] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices.* Springer Series in Statistics, 2010.